

ANALISA POLA DATA HASIL PEMBANGUNAN KABUPATEN MALANG MENGGUNAKAN METODE ASSOCIATION RULE

Dewi Sibagariang¹⁾, Karina Auliasari²⁾

^{1,2)}Jurusan Teknik Informatika, Institut Teknologi Nasional Malang
Jalan Raya Karanglo Km 2, Malang, Telp. (0341) 417634

Email : karina.auliasari86@gmail.com

Abstract

Data of development results in an area divided into several sectors. Each sector has a commodity, government use this data to determine potential commodity in small coverage area. This paper was based on our research use association rule method, as we know this method commonly used in data mining to discover pattern from huge data. Apriori is an algorithm that is implemented on application in this research, this algorithm is used to generate strong association information (strong linkage) between commodities in each sector. Support, confidence values and relationship between each commodities in 33 districts Kabupaten Malang displayed by application. From test result showed that more higher value of confidence and support make the strong relationships between commodity value. Minimum limit value can not support more than 33, because most transaction data which is calculated from the total number of 33 districts in Malang.

Intisari

Data hasil pembangunan di suatu area dibagi-bagi menjadi beberapa sektor. Tiap sektor memiliki komoditas, pemerintah menggunakan data ini untuk menentukan komoditas yang paling berpotensi di suatu daerah. Paper ini disusun berdasarkan pada penelitian menggunakan metode association rule, sebagaimana kita tahu metode ini umum digunakan pada data mining untuk menemukan suatu pola dari data yang berkapasitas besar. Algoritma apriori diimplementasikan pada aplikasi pada penelitian ini, algoritma ini digunakan untuk menghasilkan informasi keterkaitan yang kuat antar komoditas di tiap sektor. Nilai support, confidence dan hubungan antara tiap komoditas pada 33 kecamatan di Kabupaten Malang ditampilkan oleh aplikasi. Dari hasil pengujian menunjukkan bahwa semakin tinggi nilai confidence dan support maka semakin kuat nilai hubungan antar komoditi. Nilai batasan minimum support tidak bisa lebih dari 33, karena data transaksi terbanyak yang dihitung adalah dari seluruh jumlah kecamatan yang ada pada 33 kecamatan di Kabupaten Malang.

Keywords : *Development result, Data Mining, Association Rule, Apriori Algorithm*

PENDAHULUAN

Pemerintah daerah dan masyarakat bekerja sama mengelola sumber daya yang ada untuk meningkatkan pertumbuhan ekonomi dan menciptakan suatu lapangan kerja baru. Tujuan dari usaha pembangunan ekonomi selain menciptakan pertumbuhan yang setinggi-tingginya, harus pula menghapus atau mengurangi tingkat kemiskinan, ketimpangan pendapatan dan tingkat pengangguran. Kebijakan yang diambil dalam proses pembangunan daerah didasarkan pada empat faktor, yaitu kekayaan sumber daya alam, posisi geografis, potensi lahan dan potensi sumber daya manusia. Faktor-faktor tersebut mengarahkan pemerintah daerah untuk mendorong pembangunan industri, pengembangan lahan pertanian dan perkebunan, dan pengelolaan hasil perternakan dengan tepat.

Data hasil pembangunan memiliki kapasitas data yang besar dengan berbagai variasinya. Hal ini menimbulkan permasalahan dalam proses pengambilan kebijakan terkait dengan pengalokasian dana pada tiap sektor pembangunan di suatu daerah. Peran teknologi diimbangi dengan adanya sistem informasi diperlukan untuk menganalisa informasi yang dihasilkan. Oleh karena itu, diperlukan analisa terhadap pola data hasil pembangunan di setiap daerah sehingga dapat diketahui sektor dan komoditi yang berpotensi. Dengan demikian hasil analisa yang akurat dan cepat menunjang pemerintah daerah untuk memutuskan pengalokasian dana tepat sasaran.

Pencarian pola dari data yang berskala besar berkaitan dengan data mining. Data mining sendiri adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006). *Association rule* adalah salah satu teknik dalam data mining pola dari suatu kumpulan data (Kantardzic, 2003). Metode *association rule* menghasilkan aturan tertentu menghubungkan data yang satu dengan data yang lain. Untuk mencari keterkaitan data dari suatu kumpulan data, harus dicari terlebih dahulu *frequent*

itemset yaitu sekumpulan *item* yang sering muncul bersamaan. Setelah semua pola *frequent itemset* terbentuk, barulah dicari aturan asosiatif yang memenuhi tertentu.

Data hasil pembangunan di Kabupaten Malang dibagi menjadi empat bidang yaitu, geografis, perkembangan ekonomi makro, pembiayaan pembangunan, dan hasil pembangunan secara sektoral Badan (Perencanaan Kabupaten Malang, 2006). Data tersebut dalam penelitian ini dianalisis menggunakan metode *association rule* untuk menghasilkan informasi mengenai sektor pembangunan dan komoditi yang paling berpotensi pada setiap kecamatan di Kabupaten Malang.

TINJAUAN PUSTAKA

1. ASSOCIATION RULE

Aturan asosiasi yang berbentuk “if ... then ...” atau “jika ... maka ...” merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi. Dengan kata lain *Association rules* (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang “jika-maka”. Karena awalnya berasal dari studi tentang database transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan *Market Basket Analysis* atau aturan yang menyatakan asosiasi antara beberapa atribut (Han et al, 2006).

Association rule mining adalah teknik mining untuk menemukan aturan asosiatif antara suatu kombinasi *item* (Kantardzic, 2003). Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* yaitu persentase kombinasi *item* tersebut dalam *database* dan *confidence* yaitu kuatnya hubungan antar *item* dalam aturan asosiatif.

Analisis asosiasi dikenal juga sebagai salah satu metode *data mining* yang menjadi dasar dari berbagai metode data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien (Susanto, dkk, 2010). Penting

tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi *item* tersebut. dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar *item* dalam aturan assosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

2. ALGORITMA APRIORI

Algoritma apriori termaksud jenis aturan asosiasi pada *data mining*. Analisis asosiasi adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Algoritma yang paling populer dikenal sebagai Apriori dengan paradigma *generate and test*, yaitu pembuatan kandidat kombinasi *item* yang mungkin berdasar aturan tertentu lalu diuji apakah kombinasi *item* tersebut memenuhi syarat *support minimum* (Santosa,2007).

Kombinasi *item* yang memenuhi syarat tersebut disebut *frequent itemset*, yang nantinya dipakai untuk membuat aturan-aturan yang memenuhi syarat *confidence minimum*. Suatu aturan asosiatif dapat diketahui dengan dua parameter, yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah presentase kombinasi *item* tersebut dalam *database*, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-*item* dalam aturan asosiasi (Lutfi, dkk, 2009).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

1. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah *item* diperoleh dengan rumus sebagai berikut.

$$Support(A) = \frac{\sum(A \cup B)}{n} \dots\dots\dots(2.1)$$

Sementara itu, nilai *support* dari 2 *item* diperoleh dari rumus 2 sebagai berikut.

$$Support(A,B) = \frac{P(A \cap B)}{P(A \cup B)} \dots\dots\dots(2.2)$$

$$Support(A,B) = \frac{\sum(A \cup B)}{\sum A} \dots\dots\dots(2.3)$$

Keterangan :

A dan B = item atau komoditi

2. Pembentukan Aturan Asosiasi

Sebuah set dari item disebut juga *itemset*. Sebuah *itemset* yang mengandung *k items* adalah *k itemset*. Set {gula, kopi} adalah 2-*itemset*. Jumlah kejadian munculnya *itemset* adalah jumlah transaksi yang mengandung *itemset* tersebut, atau disebut *frequency*, *support count*, atau *count* dari *itemset*. Sebuah *itemset* akan memenuhi *minimum support* (*min-sup*) jika jumlah kejadian munculnya *itemset* tersebut lebih besar atau sama dengan *min-sup* yang telah ditentukan. Jika salah satu *itemset* memenuhi *support*, maka *itemset* tersebut disebut juga *frequent itemset*. Suatu set dari *frequent k-itemset* pada umumnya di notasikan dengan *L(k)*.

Langkah – langkah untuk melakukan perhitungan *association rule* terdiri dari :

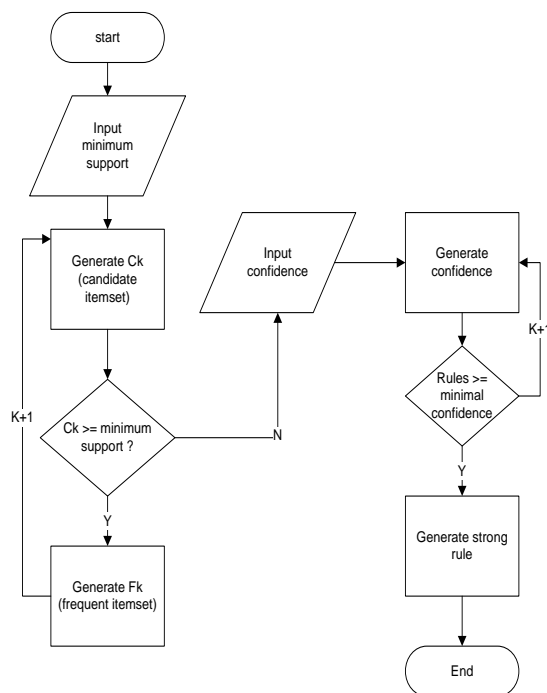
1. Temukan *frequent itemset* (ϕ), yaitu suatu item set yang memenuhi *minimum support* dan *confidence* yang telah ditentukan.
2. Menetapkan semua *frequent itemset* yaitu *itemset* yang memiliki frekuensi *itemset* minimal sebesar bilangan ϕ yang telah ditetapkan sebelumnya.
3. Gunakan *frequent itemset* tersebut untuk menghasilkan *association rule*.

PEMBAHASAN

1. IMPLEMENTASI ALGORITMA

Algoritma apriori adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut dengan istilah *minimum support*. *Minimum support* adalah batas minimum transaksi. Jika jumlah transaksi kurang dari *minimum support* maka item atau kombinasi item tidak akan diikutkan perhitungan selanjutnya. Algoritma ini akan menghasilkan kandidat baru dari *k-*

itemset dari *frequent itemset* dan menghitung nilai *support k-itemset* tersebut. *Itemset* yang memiliki nilai *support* di bawah dari *minimum support* akan dihapus. Penghitungan ini juga tidak perlu melakukan perulangan *scanning* pada *database* untuk menghitung *confidence*, cukup dengan mengambil *itemset* dari hasil *support*. Proses pada algoritma *apriori* dapat dilihat pada diagram alir pada Gambar 1.



Gambar 3.4 Diagram Alir Apriori

Algoritma akan berhenti ketika tidak ada lagi *frequent itemset* baru yang dihasilkan. *Rule* yang dihasilkan berasal dari *frequent itemset*, sehingga dalam menghitung *rule* menggunakan *confidence*, tidak perlu lagi menghitung *support* karena semua calon *rules* yang dihasilkan telah memenuhi *minimum support* sesuai yang ditentukan. Contoh dari proses mining untuk mengetahui hubungan jumlah produksi lima komoditi disektor peternakan dari lima kecamatan. Misalkan data seperti pada Tabel 1. Pada hasil Tabel 1 dapat dibentuk dalam tabular data transaksi yang akan tampak seperti Tabel 2. Dari data transaksi dengan format

tabular tersebut didapat kandidat pertama (C1) seperti pada Tabel 3. Pada kandidat pertama C1 didapat hasil seperti pada Tabel 4. Dari Tabel 4 didapat kandidat kedua (C2) seperti pada Tabel 5. Pada kandidat kedua C2 dapat ditetapkan *minimum support* = 1, maka kandidat yang kurang dari nilai 1 maka akan di hapus. Sehingga didapat hasil seperti pada Tabel 6 *minimum support* 1 (L2).

Tabel 1 Tabel Peternakan

Kode (id_kecamatan)	Item (komoditi) yang dimiliki
01	T1, T3, T4, T5
02	T3, T4, T5
03	T1, T2, T3, T5
04	T1, T2, T3, T4
05	T1, T4, T5

Tabel 2 Format Tabular Data
Transaksi Peternakan

Transaksi	T1	T2	T3	T4	T5
01	1	0	1	1	1
02	0	0	1	1	1
03	1	1	1	0	1
04	1	1	1	1	0
05	1	0	0	1	1

Tabel 3 Kandidat pertama (C1)

Itemset	Support count
T1	4
T2	2
T3	4
T4	4
T5	4

Tabel 4 Hasil pertama
(L1)

Itemset	Support count
T1	4
T2	2
T3	4
T4	4
T5	4

Tabel 5 Kandidat kedua
(C2)

Itemset	Support count
T1, T2	2

T1, T3	3
T1, T4	3
T1, T5	3
T2, T3	2
T2, T4	1
T2, T5	1
T3, T4	3
T3, T5	3
T4, T5	3

Tabel 6 Hasil setelah *minimum support* ditetapkan (L2)

Itemset	Support count
T1, T2	2
T1, T3	3
T1, T4	3
T1, T5	3
T2, T3	2
T3, T4	3
T3, T5	3
T4, T5	3

Dari Tabel 6 didapat kandidat ketiga (C3) dengan kombinasi terhadap tiga itemset dapat dilihat seperti pada Tabel 7.

Tabel 7 Kandidat ketiga (C3)

Itemset	Support count
T1, T2, T3	2
T1, T2, T4	1
T1, T2, T5	1
T1, T3, T4	2
T1, T4, T5	2
T2, T3, T4	1
T2, T3, T5	1
T3, T4, T5	2

Dari data-data tersebut dapat diambil *association rule* sebagai berikut :

1. T1, T2, T3
 $Support = \text{jumlah} \{ T1, T2, T3 \} / \text{total transaksi} = 2/5 = 40\%$
 $Confidence = \text{transaksi} \{ T1, T2, T3 \} / \text{jumlah T1} = 2/4 = 50\%$
2. T1, T2, T4
 $Support = \text{jumlah} \{ T1, T2, T4 \} / \text{total transaksi} = 1/5 = 20\%$
 $Confidence = \text{transaksi} \{ T1, T2, T4 \} / \text{jumlah T1} = 1/4 = 25\%$
3. T1, T2, T5
 $Support = \text{jumlah} \{ T1, T2, T5 \} / \text{total transaksi} = 1/5 = 20\%$

$Confidence = \text{transaksi} \{ T1, T2, T5 \} / \text{jumlah T1} = 1/4 = 25\%$

4. T1, T3, T4

$Support = \text{jumlah} \{ T1, T3, T4 \} / \text{total transaksi} = 2/5 = 40\%$

$Confidence = \text{transaksi} \{ T1, T3, T4 \} / \text{jumlah T1} = 2/4 = 50\%$

5. T1, T4, T5

$Support = \text{jumlah} \{ T1, T4, T5 \} / \text{total transaksi} = 2/5 = 40\%$

$Confidence = \text{transaksi} \{ T1, T4, T5 \} / \text{jumlah T1} = 2/4 = 50\%$

6. T2, T3, T4

$Support = \text{jumlah} \{ T2, T3, T4 \} / \text{total transaksi} = 1/5 = 20\%$

$Confidence = \text{transaksi} \{ T2, T3, T4 \} / \text{jumlah T2} = 1/2 = 50\%$

7. T2, T3, T5

$Support = \text{jumlah} \{ T2, T3, T5 \} / \text{total transaksi} = 1/5 = 20\%$

$Confidence = \text{transaksi} \{ T2, T3, T5 \} / \text{jumlah P1} = 1/2 = 50\%$

8. T3, T4, T5

$Support = \text{jumlah} \{ T3, T4, T5 \} / \text{total transaksi} = 2/5 = 40\%$

$Confidence = \text{transaksi} \{ T3, T4, T5 \} / \text{jumlah P1} = 2/4 = 50\%$

Keterangan :

Hasil perhitungan *support* dan *confidence* dapat dijadikan persen (%) dengan cara $\text{bilangan} \times 100$.

Setelah proses *mining* akan disajikan hasil dari *data mining* berupa tabel hubungan kekuatan dengan nilai *support* dan *confidence* masing-masing atribut serta *threshold* yang digunakan. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut.

2. HASIL PENGUJIAN

Proses perhitungan menggunakan algoritma apriori pada sistem menghasilkan adalah aturan akhir yang berisi keterkaitan antar komoditi. Hasil dari proses ini ditunjukkan pada Gambar 2. Pada hasil pengujian dapat dilihat pada Gambar 2 yaitu dapat memilih berdasarkan komoditi dan nilai. Dalam aturan sudah terdapat nilai *support*, *confidence* dan $\text{support} \times \text{confidence}$ yang merupakan nilai dari *strong rule*.

Nilai *support* dari suatu komoditi merupakan nilai dari seluruh data yang ada dalam database yang mengandung *itemset* (komoditi) tersebut. Nilai *support* 60 artinya adalah kemunculan suatu *itemset* (komoditi) adalah 20 dari seluruh data yang ada pada database.

FINAL ASSOCIATION RULE					
RULE					
Item	aturan	support	confidence	supportXconfide	
P1, P2	PADI SAWAH >> PADI LADANG	60	60	3673	
P1, P3	PADI SAWAH >> JAGUNG SAWAH	93	93	8824	
P1, P4	PADI SAWAH >> JAGUNG LADANG	90	90	8264	
P1, P5	PADI SAWAH >> UBI KAYU	90	90	8264	
P1, P6	PADI SAWAH >> UBI JALAP	75	75	5739	
P1, P7	PADI SAWAH >> KACANG TANAH	78	78	6207	
P1, P9	PADI SAWAH >> CENGKEH RAKYAT	60	60	3673	
P1, P11	PADI SAWAH >> KOPI ROBUSTA RAKYAT	93	93	8824	
P1, P12	PADI SAWAH >> TEBU RAKYAT	96	96	9403	
P1, P13	PADI SAWAH >> KELAPA RAKYAT	93	93	8824	
P1, P14	PADI SAWAH >> KAPUR RANDU RAKYAT	48	48	2560	
P1, P15	PADI SAWAH >> TEMBAKAU	36	36	1322	
P2, P3	PADI LADANG >> JAGUNG SAWAH	60	100	6060	
P2, P4	PADI LADANG >> JAGUNG LADANG	60	100	6060	

Gambar 2 Hasil *rule* yang terbentuk

Nilai *confidence* merupakan tingkat kepastian dari kebenaran pada *rule* yang terbentuk. *Rule* yang dicari merupakan nilai yang mempunyai kebenaran yang tinggi atau yang sering muncul.

Rule PADI SAWAH>>JAGUNG LADANG dan PADI SAWAH >> UBI KAYU dengan nilai yang sama yaitu *support* 90 memiliki nilai *confidence* 90 dan *supportXconfidence* 8264 artinya kemunculan dari PADI SAWAH >> JAGUNG LADANG dan PADI SAWAH >> UBI KAYU adalah 90 dari database.

Rule PADI LADANG>>JAGUNG SAWAH dan PADI LADANG>>JAGUNG LADANG dengan nilai sama yaitu *support* 60 memiliki nilai *confidence* 100 dan *supportXconfidence* 6060 artinya kemunculan dari PADI LADANG>>JAGUNG SAWAH dan PADI LADANG>>JAGUNG LADANG adalah 60 dari database.

Semakin tinggi nilai *support* dan *confidence* maka akan semakin kuat aturan (*strong rule*) yang dihasilkan. Pada *rule* yang terbentuk diatas memiliki nilai *confidence* yang berbeda apabila dibalik seperti PADI LADANG>>JAGUNG

SAWAH dengan JAGUNG SAWAH>>PADI LADANG karena nilai *confidence* tergantung dari *support count* suatu *itemset* (komoditi). *Support count* PADI LADANG berbeda dengan JAGUNG SAWAH yang muncul disaat bersamaan dinyatakan dengan *support* PADI LADANG dan JAGUNG SAWAH.

KESIMPULAN

Dari hasil pengujian dapat diketahui bahwa komoditi yang sering muncul dari seluruh kecamatan yang diuji maka komoditi tersebut merupakan komoditi paling berpotensi. Penerapan algoritma apriori pada aplikasi digunakan untuk menampilkan informasi tingkat jumlah produksi yang dihasilkan oleh setiap sektor pembangunan. Informasi yang dihasilkan berupa nilai *support* dan *confidence* serta keterkaitan antara tingkat jumlah produksi komoditi dengan kecamatan yang ada di Kabupaten Malang.

Pengukuran validitas algoritma menggunakan dua perhitungan untuk menunjukkan hasil yang sama dalam perhitungannya, sehingga aplikasi dapat dikatakan valid dan akurat. Data komoditi yang diproses pada *data mining* meliputi data jenis komoditi, data tahun, data kecamatan, dan jumlah produksi komoditi. Hasil dari proses *data mining* ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi jumlah produksi pada setiap jenis komoditi di sektor khususnya faktor dalam nilai perhitungan jumlah produksi per tahun. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar komoditi.

DAFTAR PUSTAKA

- [1]Badan Perencanaan Kab. Malang, 2006, "Hasil-hasil Pembangunan Kab. Malang Thn. 2004 s/d 30 Juni 2006"; Badan Perencanaan Kabupaten Malang; Malang.
- [2]Han, J. ; Kamber,Micheline, 2006; *Second Edition Data Mining Concepts and Techniques*; MORGAN

- KAUFMANN PUBLISHER; San Francisco.
- [3]Kantardzic, M.; 2003; *Data Mining: Concepts, Models, Methods, and Algorithms*; John Wiley & Sons; New Jersey.
- [4]Laboratorium Sistem Informasi; 2011; Modul Database; Institut Teknologi Nasional; Malang.
- [5]Lutfi, E.T. ; Kusrini; 2009; *Algoritma Data Mining*; ANDI OFFSET; Yogyakarta.
- [6]Pramudiono, I.; 2006; Apa Itu Data Mining?;
<http://datamining.japati.net/>
- [7]Susanto, S. ; Suryadi, D.; 2010; Pengantar DATA MINING Menggali Pengetahuan dari Bongkahan Data; ANDI OFFSET; Yogyakarta.
- [8]Santosa, B., 2007; *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu; Yogyakarta.